

Fuzzy-based Propagation of Prior Knowledge to Improve Large-Scale Image Analysis Pipelines

Johannes Stegmaier* Ralf Mikut*

Abstract

Many automatically analyzable scientific questions are well-posed and offer a variety of information about the expected outcome *a priori*. Although often being neglected, this prior knowledge can be systematically exploited to make automated analysis operations sensitive to a desired phenomenon or to evaluate extracted content with respect to this prior knowledge. For instance, the performance of processing operators can be greatly enhanced by a more focused detection strategy and the direct information about the ambiguity inherent in the extracted data. We present a new concept for the estimation and propagation of uncertainty involved in image analysis operators. This allows using simple processing operators that are suitable for analyzing large-scale spatiotemporal (3D+t) microscopy images without compromising the result quality. On the foundation of fuzzy set theory, we transform available prior knowledge into a mathematical representation and extensively use it to enhance the result quality of various processing operators. All presented concepts are illustrated on a typical bioimage analysis pipeline comprised of seed point detection, segmentation, multiview fusion and tracking. Furthermore, the functionality of the proposed approach is validated on a comprehensive simulated 3D+t benchmark data set that mimics embryonic development and on large-scale light-sheet microscopy data of a zebrafish embryo. The general concept introduced in this contribution represents a new approach to efficiently exploit prior knowledge to improve the result quality of image analysis pipelines. Especially, the automated analysis of terabyte-scale microscopy data will benefit from sophisticated and efficient algorithms that enable a quantitative and fast readout. The generality of the concept, however, makes it also applicable to practically any other field with processing strategies that are arranged as linear pipelines.

1 Background

Available prior knowledge is often not sufficiently considered by automatic processing pipelines and a great amount of potentially useful extra information re-

*Institute for Applied Computer Science, Karlsruhe Institute of Technology, Karlsruhe, Germany

mains unused. Particularly in the domain of image processing and image analysis, the visual analysis of acquired image data offers a large repository of usable *a priori* information that can often easily be verbalized by experts of the respective application field. Examples of the successful incorporation of prior knowledge are, *e.g.*, the approaches described in [1, 13] that make use of information about the expected object number as well as their associated physical size in order to adjust and improve seed point detection algorithms. Analogously, properties like size, shape, geometry, intensity distributions and the like can be used to improve the algorithmic performance of image segmentation algorithms [21, 24, 29]. Such prior knowledge is often embedded into the algorithm via shape penalization terms that are appended to the energy functional of graph-cut [29, 54] or a level-set segmentation [28] or by generalized Hough transforms that can detect arbitrary shapes [4]. Object properties like size, shape and movement dynamics can also be used to formulate efficient correction heuristics for object tracking algorithms [2, 5, 41].

A great but often underestimated potential for algorithmic improvements lies in the estimation of uncertainties of the automatically produced results and should ideally be considered by subsequent processing steps [40]. On the pixel level, this uncertainty can be used to assess the information quality of a single pixel due to sensor imperfections or temperature dependence [31, 40]. Furthermore, the localization uncertainty of geometric features such as corners, centroids, edges and lines in images has been assessed in [3, 15, 16, 36]. An approach to evaluate the quality of image registration algorithms was presented in [27]. Besides many applications from the field of quality quantification, a part of research focuses on uncertainty quantification in areas such as face recognition and other biometric technologies [9, 10, 14], the tracking of shapes in ultrasound images [57] or to evaluate the impact of noisy measurements on the validity of diagnosis results [33]. An uncertainty formulation based on fuzzy set theory has been employed to perform pixel- or object-based classification tasks [11, 39, 52]. A further possibility to exploit the uncertainty information is to optimize parameter values of a respective operator in a feedback fashion such that the outcome minimizes a previously defined optimization criterion as demonstrated in [22, 23]. Another example is the improvement of a graph-based watershed implementation, where uncertainties are used to assess the influence of individual edges on the final segmentation outcome [50].

Hitherto, however, a uniform approach to systematically transform, embed and use the available prior knowledge to improve both existing and new algorithms is missing. Although the sequential arrangement of processing operators is a broadly used concept in image analysis, results propagated through the pipelines are mostly not assessed by the individual pipeline components with respect to their result quality. Thus, errors made in early processing steps tend to accumulate and may negatively affect the final result quality. Additionally, many existing methods for processing tasks such as seed point detection, segmentation and tracking are often not directly applicable to large-scale 3D+t data sets due to enormous memory or computation time demands.

Throughout the present contribution, uncertainty is considered as the imper-

fect knowledge about the validity of a piece of extracted information produced by the respective image analysis operators and is used to derive efficient improvement heuristics to enhance image analysis pipelines [12, 47]. We present and apply a new concept for the estimation and propagation of uncertainty involved in image analysis operators that improves the result quality awareness of each processing operator and can be used to trigger adapted processing strategies. On the foundation of our previous work [47], we use fuzzy set theory to transform available prior knowledge into a mathematical representation and extensively use it to enhance the performance of image analysis operators by data filtering, uncertainty propagation and explicit exploitation of information uncertainty for result improvements. In particular, we extend an exemplary image analysis pipeline comprised of seed point detection, segmentation, multiview fusion and tracking with uncertainty handling. After introducing the general concept, we demonstrate how simple processing operators can be extended with uncertainty handling to improve large-scale analyses of 3D+t microscopy images. All methods are quantitatively validated on a comprehensive simulated validation benchmark data set that mimics embryonic development and is inspired by epiboly movements of zebrafish embryos. Moreover, we show qualitative results obtained with the presented framework on large-scale light-sheet microscopy data of developing zebrafish embryos.

2 Methods

2.1 Uncertainty Propagation in Image Analysis Pipelines

2.1.1 The Image Analysis Pipeline Concept

Most image analysis pipelines make use of multiple processing operators that are arranged as a linear processing pipeline and perform specialized tasks, such as improving or transforming the image signal, or to extract information from the images (Fig. 1). The N_{op} sequentially connected operators receive either a raw or processed version of the input image (denoted by \mathbf{I}_i), extracted features (denoted by \mathcal{X}_i) or both from their preceding processing operator with $i \in \{1, \dots, N_{\text{op}}\}$ being the ID of the operator. The output set \mathcal{X}_i of processing operator i is

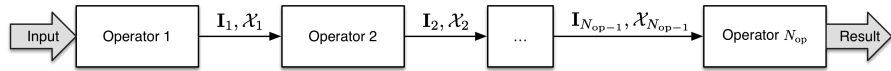


Figure 1: General image analysis pipeline comprised of N_{op} sequentially arranged processing operators. Each operator directly depends on the quality of the input images (\mathbf{I}_*) or features (\mathcal{X}_*) provided by its predecessor (adapted from [47]).

an $(N_i \times N_{\text{f},i})$ matrix with N_i data tuples and $N_{\text{f},i}$ features. For processing

operators without any feature output, \mathcal{X}_i is an empty matrix and only the processed image is passed to the next operator.

2.1.2 Identification of Suitable Prior Knowledge

Prior knowledge can be obtained from literature, expert knowledge, experimental evidence or knowledge databases. Visual analysis of acquired data often allows experts to easily identify recurring patterns, intensity properties or the appearance of objects contained in the images that can be described in natural language. An exemplary overview of such prior information derived from microscopy images is summarized in Tab. 1. The prior information are listed in bottom-up order, *i.e.*, from the acquisition stage over the content of a single image through to the spatio-temporal comparison between time series of images. Naturally, the listing is not exhaustive and the suitable features have to be carefully selected to match the underlying image material and analysis problem. In the following sections, the presented natural language expressions will be used to transform the prior knowledge of different sources to a consistent mathematical representation using the concept of fuzzy sets.

2.1.3 Quantifying Prior Knowledge using Fuzzy Set Membership Functions

To transform the prior knowledge presented in Tab. 1 into a mathematical representation, we make use of fuzzy sets that have been introduced by Zadeh in 1965 [56]. Analogous to the characteristic function of a classical set, a fuzzy set \mathcal{A} can be defined by its associated membership function (MBF) $\mu_{\mathcal{A}} : \mathcal{X} \mapsto [0, 1]$ that maps each element of a universe of discourse \mathcal{X} to a value in the range $[0, 1]$ [56]. This assigned value in turn directly reflects the fuzzy set membership degree (FSMD) of the respective element to the fuzzy set \mathcal{A} . The special cases $\mu_{\mathcal{A}}(x) = 1$ and $\mu_{\mathcal{A}}(x) = 0$ indicate that x is fully included or not part of the fuzzy set \mathcal{A} , respectively [8]. The most common membership functions used in practice are trapezoidal membership functions, which can be parameterized to model singletons, triangular and rectangular MBFs. A trapezoidal membership function can be formulated as

$$\mu_{\mathcal{A}}(x, \boldsymbol{\theta}) = \max \left(\min \left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c} \right), 0 \right), \quad (1)$$

with the parameter vector $\boldsymbol{\theta} = (a, b, c, d)^{\top}$ that is used to control the start and end points of the respective transition regions. Here, we make use of a standard partition, *i.e.*, maximally two neighboring fuzzy sets overlap and have non-zero membership values for a certain value of x and the respective membership degrees for any input value of x sum up to 1 (Fig. 2).

Considering, for instance, the detection of a specific kind of object it might make sense to use the size feature of the object as an indicator of its appropriateness. As described in [44], the linguistic terms could in this case for example be determined by five possible outcomes, where the extracted feature ...

Table 1: Prior knowledge for 3D+t image analysis and exemplary natural language expressions.

Source	Description	Example
Image Acquisition	Acquisition-specific prior knowledge such as illumination conditions, detection path, image resolution, physical spacing of voxels, high-quality image regions, point spread function (PSF) or the detection path.	Image quality decreases from XY to YZ.
Intensity	Signal-dependent information like the intensity range, time-variant characteristics of objects (<i>e.g.</i> , photobleaching in fluorescence microscopy), signal-to-noise ratio and global statistical properties of the image intensity values.	Valid objects are brighter than XY.
Localization	Positional information of the objects or object properties in absolute image coordinates. Furthermore, localization of extracted properties or objects relative to each other can be used to define neighborhood relations.	Object type XY only appears close to location YZ.
Spatial Extent	Object properties such as size, volume, principal components, convex hull extents or bounding volumes.	Object type X is larger than Y but smaller than Z.
Geometry	Geometrical properties like dimensionality, symmetry, shape, proportions and relative localization of features within an object.	Object type XY has a line-like shape with a central symmetry axis.
Morphology	Combination of intensity-based and geometrical properties, <i>e.g.</i> , to link information about patterning, texture, structure and color to geometrical properties such as shape and symmetry.	Object XY is spherical, bright and has a textured surface.
Object Interaction	Characterization of between-object properties like clustering, adhesion, repulsion, division or regional density changes.	Object behavior XY rather appears in dense regions.
Spatio-Temporal Coherence	Dynamically changing quantities such as object growth, movement direction, speed, object appearance and disappearance.	Object moves maximally XY pixels between two subsequent frames.

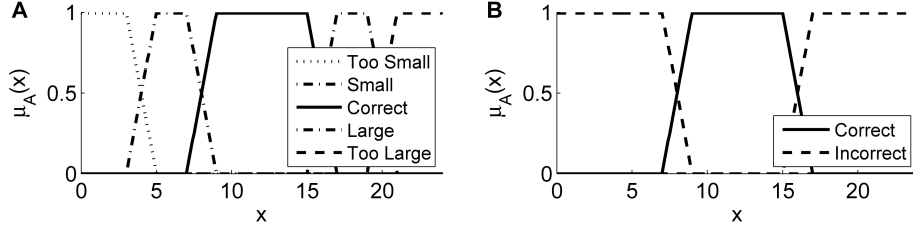


Figure 2: Different possibilities to partition the input space of a feature x using trapezoidal membership functions. In (A), each of the linguistic terms has a separate fuzzy set and (B) shows a reduced version with only two fuzzy sets that correspond to the desired class and its complement. In (B), different possibilities to summarize the correct objects arise. Besides restricting the class to the correct set as done in (B), the correct fuzzy set could be extended by the potentially useful classes (*Small* and *Large*). However, the appropriate formulation has to be chosen application dependent.

1. ... perfectly matches the expected value (*Correct*).
2. ... is smaller than expected but might be useful (*Small*).
3. ... is larger than expected but might contain useful information (*Large*).
4. ... is too small and not useful (*Too Small*, *e.g.*, noise or artifacts).
5. ... is too large and not useful (*Too Large*, *e.g.*, segments in background regions).

Available prior knowledge can be used to determine the parameterization of the associated fuzzy sets and an exemplary standard partition is shown in Fig. 2A. If only one outcome of the operators is of importance (*e.g.*, Case 1 in the above-mentioned example), it is possible to use only one linguistic term and to aggregate all other cases by its complement (Fig. 2B). We use $\mu_{\mathcal{A}_{i fl}} : \mathbb{R} \rightarrow [0, 1]$ to denote the fuzzy set membership function for image analysis operator i , feature $f \in \{1, \dots, N_{f,i}\}$ and linguistic term $l \in \{1, \dots, N_l\}$. Thus, the n -th data tuple produced by operator i obtains the FSMD value $\mu_{\mathcal{A}_{i fl}}(x_i[n, f])$ to the fuzzy set $\mathcal{A}_{i fl}$ for each feature f and each linguistic term l .

2.1.4 Combination of Fuzzy Set Membership Functions

Fuzzy set membership degree values of multiple features that characterize a linguistic term (*e.g.*, if an object of interest is bright and elongated at the same time) can be combined using a fuzzy pendant to a logical conjunction [12]. A conjunction of $N_{f,i}$ fuzzy set membership functions for linguistic term l can be defined using the minimum operator:

$$\mu_{lc, \mathcal{A}_{il}}(\mathbf{x}_i[n]) = \min_{f=1, \dots, N_{f,i}} (\mu_{\mathcal{A}_{i fl}}(x_i[n, f])). \quad (2)$$

Features that should not contribute to the combined fuzzy set membership function can be disabled by setting the corresponding MBFs to the constant value 1 (identity element of the conjunction) and the complement of the combined linguistic term is simply given by $1 - \mu_{lc, \mathcal{A}_{il}}(\mathbf{x}_i[n])$ as illustrated in Fig. 2B. Compared to the multiplication-based conjunction used in [47], the minimum-based formulation in Eq. (2) is more informative, as a non-zero value directly represents the lowest FSMD value of the considered fuzzy sets.

2.1.5 Uncertainty Propagation in Image Analysis Pipelines

We use the FSMD values associated with each data tuple to perform a feed-forward propagation of the reliability of extracted data to downstream operators. For each data vector that is produced by operator i , we calculate the degree of membership to the respective fuzzy sets and append it to the feature output \mathcal{X}_i . If only a classification into correct vs. incorrect objects needs to be performed (Fig. 2B) or if a linguistic term is described by a combination of different fuzzy sets (Eq. (2)), a single FSMD value is appended per data tuple. The FSMD values can be used to perform object filtering, extended information propagation and to resolve ambiguities.

Uncertainty-based Object Rejection: The first application of the uncertainty framework is to filter the extracted output information \mathcal{X}_i produced by an operator i using thresholds $\alpha_{il} \in [0, 1]$. According to the FSMD values $\mu_{lc, \mathcal{A}_{il}}(\mathbf{x}_i[n])$ calculated for each data tuple $\mathbf{x}_i[n] \in \mathcal{X}_i$, $\mathbf{x}_i[n]$ is only passed to the next pipeline component if $\mu_{lc, \mathcal{A}_{il}}(\mathbf{x}_i[n]) \geq \alpha_{il}$ for the membership to a desired set. The reduced set which serves as input for operator $i + 1$ is denoted by $\tilde{\mathcal{X}}_i$. To keep all extracted information, the threshold is set to $\alpha_{il} = 0$. Contrary, for $\alpha_{il} = 1$, no uncertain information is passed to operator $i + 1$. Based on application specific criteria or object properties, this FSMD-based object rejection allows to easily filter out false positive detections as demonstrated in Sec. 2.2.1 and Sec. 2.2.2, respectively.

Extended Information Propagation to Compensate Operator Flaws: Second, we allow operators to fall back on information of penultimate processing steps if predecessors do not deliver good results. For instance, if an operator i fails to sufficiently extract information from its provided input data (*e.g.*, missing, merged or misshapen objects), it can inform downstream operators about these flawed results. Using a second threshold $\beta_{il} \in [\alpha_{il}, 1]$ for each operator, the FSMD level below which the information of the previous steps should be additionally propagated can be controlled. More formally this means that instead of only forwarding the α_{il} -filtered set $\tilde{\mathcal{X}}_i \subseteq \mathcal{X}_i$ produced by operator i to operator $i + 1$ a set $\Omega_i = \tilde{\mathcal{X}}_i \cup \tilde{\Omega}_{i-1}$ with $i \geq 2$ and $\Omega_1 = \tilde{\mathcal{X}}_1$ is passed through the pipeline.

$\tilde{\Omega}_{i-1}$ represents the subset of elements in Ω_{i-1} which were not successfully transferred into useful information by operator i , *i.e.*, elements $\mathbf{x}_{i-1}[n] \in \Omega_{i-1}$ that generated output $\mathbf{x}_i[n] \in \tilde{\mathcal{X}}_i$ with $\alpha_{il} \leq \mu_{lc, \mathcal{A}_{il}}(\mathbf{x}_i[n]) < \beta_{il}$. Such elements

characterize information of operator $i - 1$ that might be useful in later steps to correct flawed results of operator i . If $\beta_{il} = 1$ all information of $i - 1$ that produced an uncertain outcome is propagated to the successor $i + 1$. If $\beta_{il} = \alpha_{il}$ only the information $\tilde{\mathcal{X}}_i$ produced by operator i is propagated. In the current version of the framework, the respective processing operators are responsible for calculating Ω_i and to appropriately calculate the respective FSMD values.

This approach was successfully used to resolve tracking conflicts that originated from under-segmentation errors as described in Sec. 2.2.3.

Resolve Ambiguities using Propagated Uncertainty: In addition to filtering and propagating the information of the operators within the pipeline, the provided uncertainty information can explicitly be used by the processing operators to improve their results. Depending on the degree of uncertainty of provided information, parameters or even whole processing methods can be adapted if needed. Of course, the adaptations required by a particular algorithm cannot be generalized. However, we showcase two potential applications in the next sections, namely the fusion of redundant seed points (Sec. 2.2.1) and the correction of under-segmentation errors (Sec. 2.2.2).

The general scheme for the proposed uncertainty propagation framework is summarized in Fig. 3 and is applied to an exemplary image analysis pipeline in the next sections.

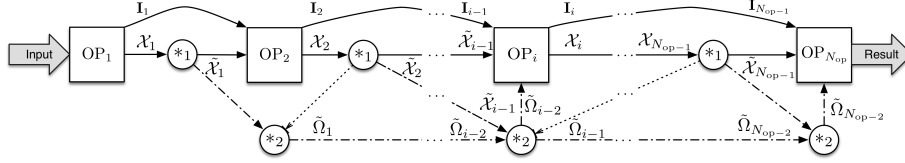


Figure 3: Extended image analysis pipeline concept. Extracted output information of each operator can be filtered according to its uncertainty (\ast_1), operators can access information produced by penultimate predecessors (\ast_2) and processing operators can specifically adjust their processing behavior based on FSMD values of extracted information. Solid lines indicate the main information flow, dash-dotted lines the propagation of previously calculated results and dotted lines emphasize influence on the selection of propagated information. In addition to the flow of extracted features (\mathcal{X}_\ast , Ω_\ast , $\tilde{\mathcal{X}}_\ast$, $\tilde{\Omega}_\ast$), the operators may pass processed images, image parts or forward the input image (\mathbf{I}_\ast) to the subsequent processing operator.

2.2 Extending and Enhancing Algorithms with Uncertainty Treatment

Based on the general concept presented in the previous section, we applied it to an exemplary image analysis pipeline comprised of seed point detection, segmentation, multiview fusion and object tracking. For each operator, FSMD values were estimated based on prior knowledge and used for algorithmic improvements where possible. We use a simulated benchmark data set that mimics 3D microscopy images containing fluorescently labeled nuclei of an artificial embryo.

The advantages of using simulated image data for validation is the possibility to have a single comprehensive data set for the validation of all pipeline components. Instead of testing each component separately on different benchmarks, this allows to uncover specific bottlenecks or error sources in the processing pipelines. Furthermore, different acquisition deficiencies such as different point-spread-functions, decreasing signal-to-noise ratios or multiview acquisition deficiencies can be simulated. The immediate availability of a reliable ground truth enables a quantitative validation without the bias observed for manually annotated benchmark data that suffers from intra- and inter-expert variability. As the simulated benchmark is close to the target application of the pipeline, namely quantitatively analyzing terabyte-scale 3D+t fluorescence microscopy images, the developed concepts and algorithms can easily be put into practice, *e.g.*, for false positive reduction of a segmentation algorithm or for segmentation-based multiview fusion [25, 43].

Details on the benchmark generation can be found in App. A and [46]. Abbreviations for the different algorithms are given in round brackets and a quantitative comparison of the result quality is provided in Sec. 3.

2.2.1 Seed Point Detection

In [48], a blob detection method based on the Laplacian-of-Gaussian (LoG) maximum intensity projection was used to localize fluorescently labeled cellular nuclei in 3D microscopy images. In brief, a 3D input image is filtered with differently scaled LoG filters with standard deviations σ matching the expected object radius r using the relation $\sigma = r/\sqrt{2}$. Subsequently, the 3D maximum projection of these LoG-filtered images is formed and local extrema are extracted from this projection image (LoGSM). Although, the proposed method worked well in many scenarios, it frequently missed objects that did not exhibit a strict local maximum due to an intensity plateau (*e.g.*, elongated objects, overexposure or discretization artifacts). To get rid of this behavior, we used the \leq -operator instead of the $<$ -operator to additionally detect non-strict local extrema (LoGNSM). However, this increased the amount of false positive detections in background regions and along elongated objects.

Detections in background regions were removed using an intensity threshold (t_{wmi}) applied on the mean intensity of a small window surrounding the potential detection. The remaining seed points were mostly located properly on the detected objects and remaining false positive detections largely originated

from objects that were detected multiple times. To combine redundant objects to a single one, a fusion approach based on hierarchical clustering was used (LoGNSM+F). The hierarchical cluster tree was computed using Ward’s minimum variance method to compute distances between clusters, *i.e.*, the within-cluster variance was minimized to obtain equally sized clusters [55]. The final clustering was obtained from the complete cluster tree using a distance-based cutoff t_{dbc} that was set to the smallest expected object radius r_{min} , to fuse close redundant detections and to prevent fusion of neighboring objects. A single detection per object was obtained by averaging the feature vectors of all detected seeds in a cluster.

As the seed detection stage usually represents one of the first analysis steps, no preceding uncertainty information was considered. To inform down-stream operators about the expected result quality, the uncertainty of the detected seed points was estimated using the window mean intensity, the maximum seed intensity and the z-position features of the extracted objects (LoGNSM+F+U). Besides discarding obvious false positive detections in background regions (intensity-based thresholds), the fuzzy sets for the z-position were adjusted such that seed detections in low contrast regions (farther away from the detection objective) had lower membership degrees to the class of correct objects than objects in the high contrast regions (closer to the detection objective). The final fuzzy set membership degree of an object to the fuzzy set of correct objects was determined using the minimum of the obtained membership degrees and was appended as a new feature to the output matrix of the seed detection algorithm. To filter false positives the forward threshold slightly above zero to $\alpha_{11} = 0.0001$ (forward threshold for Operator 1 and Linguistic Term 1). A color-coded visualization of the detected seed points and the fuzzy sets for the different features are shown in Fig. 4.

2.2.2 Segmentation

After the seed detection stage, a segmentation operator is used to extract the regions and regional properties of all detected objects from the simulated 3D image stacks, *e.g.*, using the algorithms presented in [1, 20, 48]. For demonstration purposes, we further improved an algorithm based on adaptive thresholding using Otsu’s method [35] and a watershed-based splitting of merged objects [7] (OTSUWW) as described in [48]. Therefore, propagated information from the seed detection stage and estimated FSMD values of extracted segments are used to improve the algorithmic efficiency and the segmentation quality (OTSUWW+U).

Based on the extracted statistical quantities of the benchmark images (Tab. 2), we derive the parameter vector $\boldsymbol{\theta} = (a, b, c, d)^\top$ of the trapezoidal fuzzy set membership function for each considered feature using the minimum and maximum values as a, d parameters, respectively. The remaining parameters b, c were set to the 5%-quantile and the 95%-quantile. This parameterization ensured that all values smaller or larger than the maximum values obtained a membership degree of zero and that 90% of the data range has a membership value of one

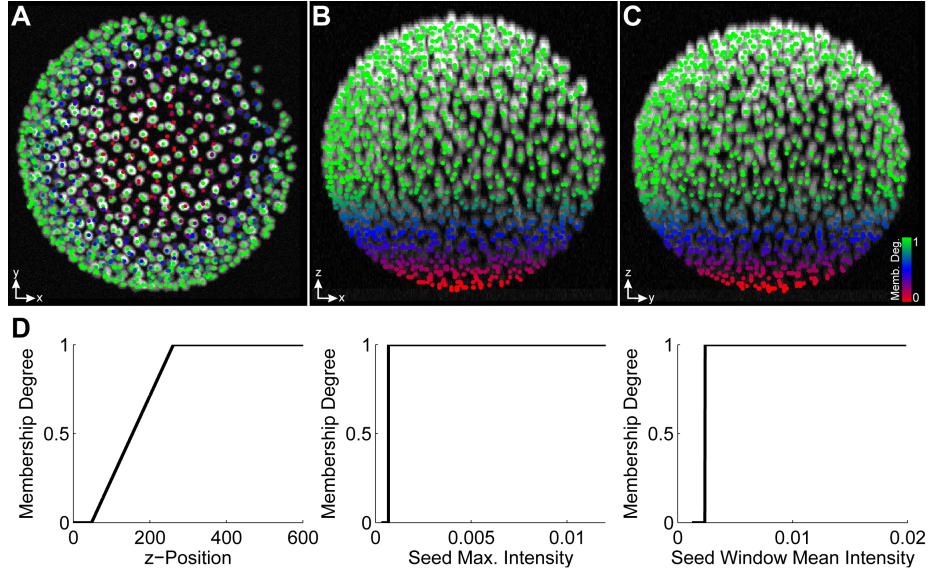


Figure 4: Maximum intensity projection of a 3D benchmark image along the Z, Y and X axis superimposed with detected seed points (A, B, C). Seed points are colored according to their FSMD to the class of a correct detection ranging from red over blue to green for low, medium and high membership degree, respectively. The fuzzy sets used for the individual features are depicted in (D) and the \min -operator was used as a fuzzy conjunction to obtain the final membership degree. The uncertainty gradient along the z-axis was introduced due to the signal attenuation at locations farther away from the detection objective and was used in later steps to resolve multiview fusion ambiguities.

assigned. Of course, this parameterization is application and data dependent and can be customized, *e.g.*, to adjust the behavior for extrema at the lower and upper spectrum of the value range. For simplicity, the focus was put on volume and size information of the objects. In the absence of ground truth data, the transition regions for the fuzzy sets can be identified by a manual analysis of objects that deviate from the expectation at the lower and the upper feature value range, *e.g.*, using software tools such as Fiji, ICY or Vaa3D [19, 37, 42].

Table 2: Statistical quantities of the benchmark data set

Feat.	Min	Max	Mean	Std.	Med.	5% qt.	95% qt.
Vol.	449	2016	993.6	247.2	990	617	1405
Width	13	31	19.9	2.6	20	15	24
Height	13	34	19.9	2.5	20	15	24
Depth	3	11	6.2	1.0	6	5	8

Minimum, maximum and quantile values were used to formulate fuzzy sets for each of the features. Individual fuzzy sets were combined using the minimum operator, to obtain a single membership degree value to the fuzzy set of being a valid object.

As depicted in Fig. 5, the shapes of the fuzzy set membership functions derived from the statistical quantities resemble the respective distribution observed in the feature histograms. We used the min-operator to combine the individual fuzzy set membership degrees to a single value, *i.e.*, the combined FSMD value directly corresponded to the membership degree of the feature that deviated the most from the specified expected range. Of course, the size criteria discussed here should only be considered as an exemplary illustration. There are various other features that can potentially be used to assess and improve segmentation results, *e.g.*, integrated intensity, edge information, local entropy, local signal-to-noise ratios (SNR), principal components, weighted centroids and many more. Furthermore, if colocalized channels are investigated, complementary information can be used to formulate more complex decision rules. In the case of a fluorescently labeled nuclei and membranes that are imaged in different channels [21, 24, 45], rules like *"each cell has exactly one nucleus"* can be formalized in the same way using the fuzzy set membership functions for a quantification of the available prior knowledge.

The identified FSMD values of segmented objects, were then used to detect under-segmentation errors produced by Otsu's method that needed be to further split to match the expected object size. This was realized in the investigated OTSUWW+U implementation using the seed points of the LoGNSM+F+U method (Sec. 2.2.1). In contrast to the previous approach (OTSUWW, [7, 48]), the watershed-based splitting technique was only applied to objects that were

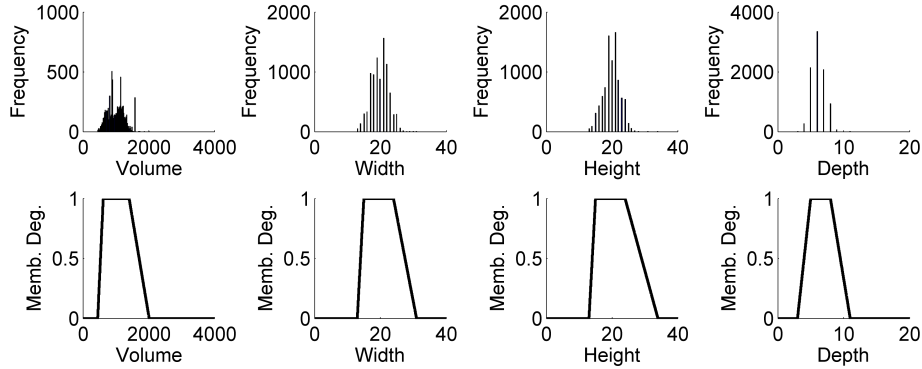


Figure 5: Feature histograms (top) and fuzzy set membership functions (bottom) for volume ($\theta_{\text{vol}} = (449, 617, 1405, 2016)^\top$), width ($\theta_{\text{w}} = (13, 15, 24, 31)^\top$), height ($\theta_{\text{h}} = (13, 15, 24, 34)^\top$) and depth ($\theta_{\text{d}} = (3, 5, 8, 11)^\top$).

known to be larger than expected. Using a parallelization strategy similar to the one discussed in [48], all segments with combined FSMD values below β_{21} (backward threshold for Operator 2 and Linguistic Term 1) that corresponded to objects larger than expected (Case 3, Sec. 2.1.3) were distributed among the available CPU cores and a seeded watershed approach was used for object splitting in each of the cropped regions of the image [7]. This approach was much faster than directly applying the watershed algorithm on the entire image (OTSUWW), due to the uncertainty-guided, locally applied processing of erroneous objects. After splitting merged objects, the connected components of the image needed to be identified once more and the uncertainty values were re-evaluated as well, to provide the updated information to the subsequent processing operators.

To further improve the results with respect to false positive detections observed for higher noise levels, segments with combined FSMD values below $\alpha_{21} = 0.1$ (forward threshold for Operator 2 and Linguistic Term 1), *i.e.*, objects that were smaller than the expected object size (Case 2 and Case 4, Sec. 2.1.3) were removed from the label images. To facilitate the implementation of the uncertainty-guided segmentation, it was realized using just a single fuzzy set for the correct class of objects (Case 1, Sec. 2.1.3) to identify objects that needed further consideration. To determine if objects with low FSMD values were smaller or larger than the expectation, a comparison to the boundaries of the expected valid range was performed (trapezoidal fuzzy set parameters b, c). Threshold values were identified using the interactive graphical user interface presented in [45].

2.2.3 Tracking

The final step of the investigated image analysis pipeline was the tracking of all detected objects, *i.e.*, to identify the correct correspondences of detected objects

in subsequent time points of the acquired images. For illustration purposes, we used a straightforward nearest-neighbor tracking approach implemented in the open-source MATLAB toolbox Gait-CAD [44]. Each object present in a frame was associated with the spatially closest object in the subsequent frame. This procedure was applied to every frame of the data set in order to get a complete linkage of all objects.

In addition to tracking the results obtained from the segmentation methods introduced in the previous section, we present an alternative approach that combines a flawed segmentation with provided seeds (OTSU+NN+U). Inspired by conservation tracking methods [41], we leave the flawed segmentation results produced by the respective algorithms unchanged and additionally provide detected seed points to the tracking algorithm instead of actually using the seed points for splitting in the image domain [47]. As shown in Tab. 4, the segmentation quality achieved by OTSU is in principle not reasonably usable without the watershed-based object splitting. Nevertheless, the tracking algorithm could be extended to decide which of the information was reliable and suitable for tracking and could optionally fall back to the provided seed points if the segmentation quality was insufficient. Therefore, we used the FSMD values provided by the segmentation stage. Using an empirically determined threshold of $\beta_{31} = 0.9$ (backward threshold for Operator 3 and Linguistic Term 1), all objects with an aggregated FSMD value lower than the threshold were not tracked with the actual segment, but with the seed points the respective segment contained. The forward threshold parameter was set to $\alpha_{31} = 0.0$ in order to report all tracking results.

3 Results and Discussion

The functionality of the proposed approaches was validated on simulated 3D benchmark images. The data sets contained images with different numbers of objects (SBDE1), different noise levels (SBDE2) and a set of 50 sequential time points with 1000 moving and interacting objects (SBDE3). The SBDE3 data set additionally included a multiview simulation, *i.e.*, at each time point, two simultaneous images from opposite direction were generated. An overview of the generated benchmark data sets is provided in Tab. 6 and a brief description on the data set generation is provided in App. A as well as Fig. 11 and Fig. 12.

3.1 Seed Point Detection Validation

To validate the proposed improvements of the LoG-based seed detection algorithm, the SBDE1 and SBDE2 benchmark data sets were used (Tab. 6) with the parameters listed in Tab. 7 and the performance measures described in App. B. The obtained values are summarized in Tab. 3, whereas each entry of the table corresponds to the arithmetic mean value of the independently obtained results on the ten benchmark images of SBDE1.

The quantitative analysis confirmed that the proposed extensions of LoGSM

Table 3: Quantitative assessment of the seed detection performance

Method	TP	FP	FN	Rec.	Prec.	F-Sc.	Dist.	Time (s)	KVox./s
LoGSM	681.1	3.3	202.8	0.77	1.00	0.87	1.60	7.23	7259.82
LoGNSM	813.7	160.0	70.2	0.91	0.84	0.87	1.64	7.66	6858.39
LoGNSM+F	811.7	4.5	72.2	0.91	0.99	0.95	1.59	7.70	6819.58
LoGNSM+F+U	812.5	4.3	71.4	0.91	0.99	0.95	1.59	9.77	6170.05

Quantitative performance assessment of the LoG-based seed detection methods. The criteria are true positives (TP), false positives (FP), false negatives (FN), recall, precision, F-Score, the distance to the reference (Dist., smaller values are better) as well as the achieved time performance measures in seconds (smaller values are better) and voxels per second (larger values are better). All values represent the arithmetic mean of the individually processed benchmark images.

could improve the algorithmic performance by up to 9.2% with respect to the F-Score. LoGSM had few false positive detections but on the other hand missed many objects due to the strict maximum detection (recall of 0.77 and precision of 1.0). The recall could be improved by 18.2% to a value of 0.91 by additionally allowing non-strict maxima (LoGNSM). However, this adaption concurrently raised the number of false positives and thus lowered the precision by 16.0% to 0.84, as objects with maximum plateaus were detected multiple times. These multi-detection errors could be successfully removed using the proposed fusion technique, which was reflected in an F-Score value of 0.95 for LoGNSM+F(+U), *i.e.*, compared to the LoGSM method, the F-Score was increased by 9.2%. Regarding the processing times, the additional effort for a redundant detection was almost negligible, as the non-strict maximum detection simply detected more seed points during the same iteration over the image. The seed point fusion was performed directly in the feature space and was therefore also insignificant compared to the preceding processing steps. For the feature set described here, using the uncertainty-based object rejection (LoGNSM+F+U) only slightly improved the results compared to directly fusing and filtering the data using the hard intensity threshold but increased the processing time by 35%. LoGNSM+F yielded almost identical results and required only 6% more processing time compared to LoGSM. Nevertheless, all objects were equipped with an uncertainty value that was propagated through the pipeline and proved to be beneficial to filter, fuse and correct the extracted data in subsequent steps. In addition, it should be noted that the processing time required for the image analysis easily exceeds the fuzzy set calculations as soon as the images get larger.

Furthermore, we tested the performance of the seed detection under different image noise conditions using the SBDE2 data set, which contained images with different settings for the additive Gaussian noise standard deviation ($\sigma_{\text{agn}} \in [0.0005, 0.01]$). Seeds from these images were extracted using the

LoGSM, LoGNSM, LoGNSM+F and LoGNSM+F+U algorithms and the intensity thresholds were determined for each of the noise levels individually using the semi-automatic graphical user interface described in [49]. For higher noise levels, the number of detections in background regions heavily increased and it became ambiguous to determine true positive detections in low-contrast regions. The manual threshold was therefore adjusted such that the false positive detections were minimized and only unambiguous seeds were considered. This continuous threshold adaption is also the reason for constant (Fig. 6A, C) or increasing precision (Fig. 6B), as it was easier to identify false positives rather than false negative detections in noisy image regions. Objects were robustly detected down to a signal-to-noise ratio of 5 (Fig. 6), which was close to the visual limit of detection [34] and emphasized the uncertainty-based improvements.

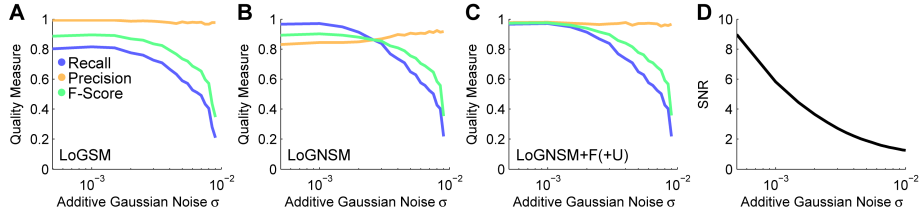


Figure 6: Assessment of the seed detection performance for the different noise levels of the SBDE2 data set. The performance measures recall, precision and F-Score are plotted versus the additive Gaussian noise level parameter σ_{agn} for LoGSM (A), LoGNSM (B) and LoGNSM+F(+U) (C). As LoGNSM+F and LoGNSM+F+U produced identical results with respect to recall, precision and F-Score, the plots are combined to a single panel (Tab. 3). The influence of the noise level on the signal-to-noise ratio of the images is plotted in (D).

3.2 Segmentation Validation

The segmentation performance was validated using the SBDE1 and the SBDE2 data sets (Tab. 6). In addition to the algorithms described in Sec. 2.2.2 (OTSU, OTSUWW, OTSUWW+U), the segmentation quality obtained by the TWANG method (see [48] for details) was added and quantitatively compared to a TWANG version (TWANG+U) that relied on the improved seed detection operator introduced in the previous section (LoGNSM+F+U). The respective parameterization and a brief description of each algorithm is provided in Tab. 8 and the validation measures are summarized in App. B and [18]. In Tab. 4, the quantitative segmentation quality results obtained on the SBDE1 data set are summarized. The Rand index (RI) value was almost identical for all algorithms and OTSU yielded the highest value. The enhanced adaptive threshold-based techniques yielded an 11.0% better Jaccard index (JI) value than TWANG+U and the best normalized sum of distances (NSD) value was obtained by OTSUWW+U. Considering the results for RI, JI and NSD, the global threshold-based techniques

(OTSU*) produced slightly more accurate results (0.3%, 5.3% and 12.0%, respectively) for the objects they were still able to resolve compared to the best results of the TWANG-based methods. Both TWANG-based methods on the other hand produced the minimal amount of topological errors with respect to split and merged objects compared to all OTSU-based methods. The number of added objects was minimal for OTSUWW+U. TWANG+U produced slightly more added objects than TWANG but efficiently detected far more objects. Thus, the F-Score values achieved by TWANG+U were further increased by 8.4% compared to TWANG, *i.e.*, TWANG+U produced the best results with the fewest topological errors (F-Score 0.9). The low amount of split and merged nuclei for TWANG originate from the single-cell extraction strategy, rather than using a global threshold as performed in the OTSU-based methods. The relatively large amount of added objects detected by the TWANG segmentation were mostly no real false positive detections, but segments where most of the extracted region intersected with the image background instead of the actual object and are thus considered as false positives.

Regarding the processing times OTSU was the fastest approach, but due to the poor quality without the uncertainty-based extension it was not really an option for a reliable analysis of the image data. TWANG and TWANG+U were 1.8 times slower than the plain OTSU method but 2.4 and 3.8 times faster than OTSUWW and OTSUWW+U, respectively. At the same time, TWANG+U was the most precise approach (F-Score 0.9). Furthermore, OTSUWW+U was 1.6 times faster than OTSUWW due to the focused object splitting and additionally produced better results due to the improved seed detection and noise reduction.

These results confirmed that the uncertainty information could be efficiently exploited to guide computationally demanding processing operators to specific locations and thus, to speed up processing operations while the result quality was preserved or even improved.

Exemplary FSMDs of the final segmentation results of different algorithms are depicted in Fig. 7 and provide a convenient visualization to instantly assess the segmentation quality and to identify potential problems of the methods even by non-experts. All algorithms suffered from the light attenuation in the axial direction. Especially, the techniques that relied on a single global intensity threshold had problems to identify the objects located in these low-contrast regions. Besides missing many objects, OTSU additionally merged many of the high intensity objects to a single large blob. Especially, in the z-direction many mergers occurred due to the lower sampling in this direction. However, these merged regions could be successfully split to a large extent using the proposed seed-based splitting techniques (OTSUWW, OTSUWW+U). As TWANG directly operated on the provided seeds, it was still able to extract most of the objects in these regions and yielded even higher recall values using the LoGNSM+F+U seed points. However, due to the low-contrast, the segmentation quality of the extracted segments in these regions was reduced. In Tab. 3, this is reflected by the increased number of added objects for TWANG and TWANG+U, which were mostly no real false positives as described above.

Table 4: Quantitative assessment of the segmentation performance

Method	RI	JI	NSD ($\times 10$)	HM	Split	Merged	Added	Missing	Rec.	Prec.	F-Score	Time (s)	KVoxel/s
OTSU	97.92	27.82	3.77	6.67	25.00	370.50	42.40	264.00	0.29	0.78	0.42	6.11	8589.46
OTSUWW	97.91	27.99	3.29	6.48	88.60	96.30	57.90	276.20	0.58	0.78	0.67	42.41	1237.72
OTSUWW+U	97.91	27.98	3.15	5.61	50.60	57.80	13.30	268.40	0.64	0.90	0.75	26.40	2035.66
TWANG	97.67	26.59	3.58	4.62	0.00	6.70	75.80	193.50	0.77	0.90	0.83	11.02	4774.83
TWANG+U	97.81	25.30	3.64	4.75	0.00	12.40	103.70	59.50	0.91	0.89	0.90	11.15	4728.84

The criteria used to compare the algorithms are the Rand index (RI), the Jaccard index (JI), the normalized sum of distances (NSD) and the Hausdorff metric (HM) as described in [18]. Additionally, the topological errors were assessed by counting split, merged, added and missing objects. Precision, recall and F-Score are based on the topological errors by considering split and added nuclei as false positives and merged and missing objects as false negatives, respectively. The achieved time performance was measured in seconds (smaller values are better) and voxels per second (larger values are better). All values represent the arithmetic mean of the individually processed benchmark images.

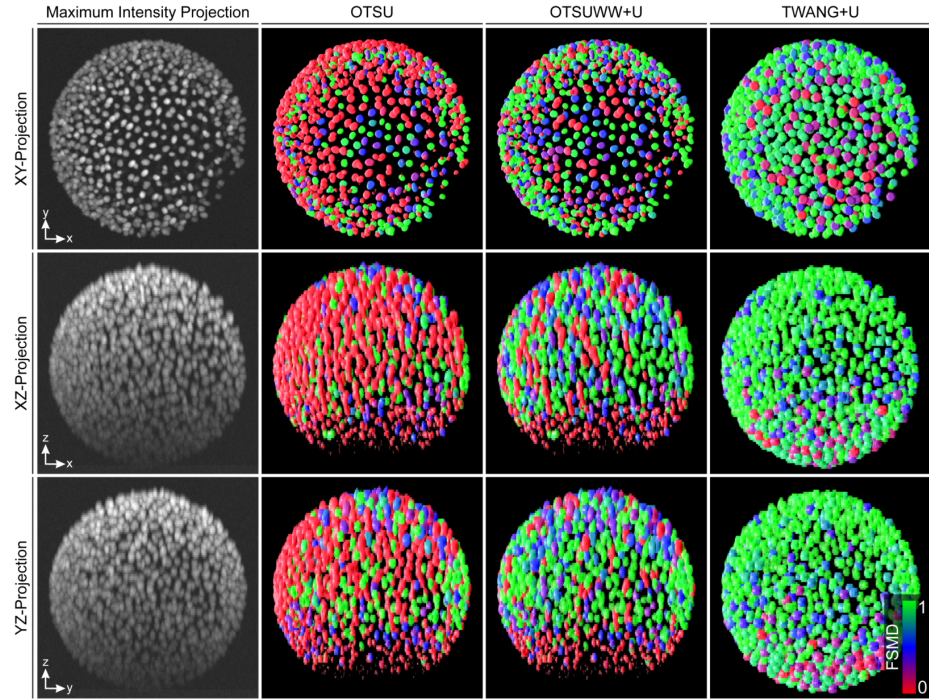


Figure 7: Maximum intensity projections of the raw image and exemplary volume renderings of the automatic segmentation results produced by OTSU, OTSUWW+U and TWANG+U from different viewpoints (XY, XZ and YZ). The FSMD of individual detected objects was estimated using the morphological criteria volume and size and was used for coloring (red over blue to green for low, medium and high membership degree, respectively).

To investigate the impact of the signal-to-noise ratio of the images to the segmentation quality, the benchmark data set **SBDE2** was processed using all five algorithms. The segmentation quality of all adaptive thresholding-based methods was heavily affected by the noise level of the images yielding poor precision and recall values even for the lowest noise levels (Fig. 8A, B). This was caused by the global threshold that on the one hand merged a lot of objects and on the other hand detected a high amount of false positive segments. The uncertainty-based method OTSUWW+U successfully preserved the increased recall of OTSUWW and at the same time substantially increased the precision to an almost perfect level for noise parameters of $\sigma_{\text{agn}} < 0.003$ (Fig. 8C). The increasing number of small segments observed for OTSU and OTSUWW could be efficiently filtered using the uncertainty-based object rejection. As TWANG heavily depended on the quality of the provided seeds, the observed curves in Fig. 8C, D show a high correlation to the seed detection performance (Fig. 6A, C) and render it as a suitable method even for higher noise levels. The improved seed detection of LoGNSM+F+U also directly affected the quality of the TWANG+U segmentation. Note that the seed points were adjusted for each of the noise levels, *i.e.*, subtle variations in the precision and recall (Fig. 8D, E) values are caused by a subjective manual threshold adaption.

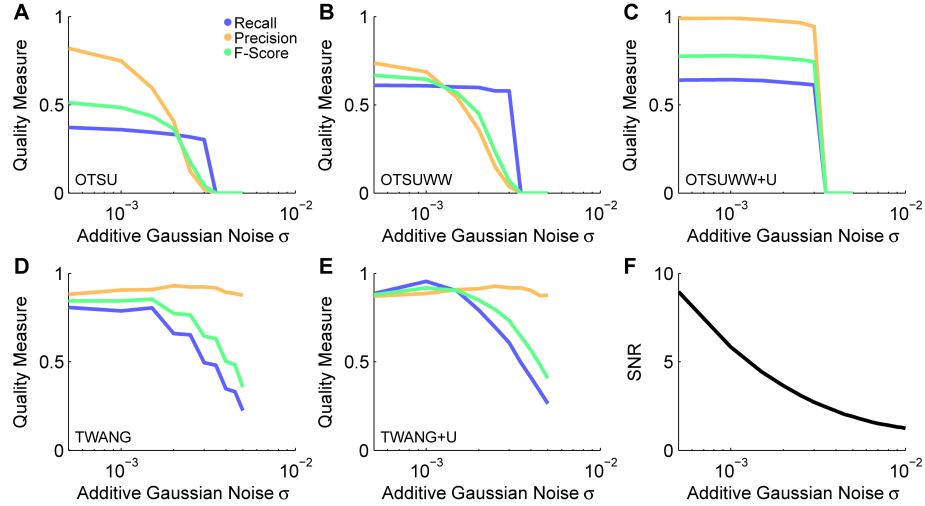


Figure 8: Performance evaluation of the segmentation methods OTSU (A), OTSUWW (B), OTSUWW+U (C), TWANG (D) and TWANG+U (E) on images of the **SBDE2** data set with different signal-to-noise ratios. The methods based on adaptive thresholding (OTSU, OTSUWW) suffered from high noise levels and produced a successively increased amount of false positive detections, which could be efficiently suppressed using the uncertainty framework-based extension (OTSUWW+U). The result quality of both TWANG versions directly correlated with the quality of the provided seed points, *i.e.*, TWANG+U benefited from the improved detection rate of LoGNSM+F+U.

3.3 Tracking Validation

The tracking validation was performed on the **SBDE3** data set, which consisted of 50 frames with two simultaneously acquired rotation images for each frame, yielding a total number of 100 frames that needed to be processed. Segmentation was performed using OTSUWW, OTSUWW+U, TWANG and TWANG+U separately for all time points and view angles. To obtain a single set of objects for each time point, segmentation results of different view angles were fused using a segment-based fusion approach described in App. C (Additional file 1). The centroids of all detected objects were then used to perform the nearest neighbor tracking (NN). In addition, the method described in Sec. 2.2.3 (OTSU+F+NN+U) was applied to the test data set and the Otsu-based threshold was applied to both rotation images independently and the resulting binary images were then fused by simply using the maximum pixel value of the two images. FSMD values were then estimated on the connected components of the fused image using the same fuzzy sets as for the segmentation step. The obtained tracking results are summarized in Tab. 5 and Fig. 9.

Both pipelines without uncertainty treatment reached the lowest tracking accuracy with respect to the tracking (TRA, see App. B) quality measure due to an increased number of missing objects (recall of 0.89 for OTSUWW+NN, and 0.88 for TWANG+NN) and a high number of false positive detections (precision of 0.82 for OTSUWW+NN). Of course, these missing objects directly correlated with the number of missing edges and explain the 222.4% (OTSUWW+NN) and 139.9% (TWANG+NN) higher amount of missing edges compared to the best scoring algorithm in this category (OTSU+F+NN+U). Furthermore, OTSUWW+NN suffered from many merged regions that also contributed to the 8.2% lower recall value compared to OTSU+F+NN+U. In contrast to this, all uncertainty-enhanced methods provided comparable results, with the best results achieved by OTSU+F+NN+U and TWANG+U+NN. Although the amount of false negatives of TWANG+U+NN was almost halved compared to TWANG+NN, the detection rate was still the largest problem of this pipeline resulting in a 21.2% higher amount of missing edges compared to OTSU+F+NN+U. On the other hand, OTSU+F+NN+U still suffered from the same under-segmentation tendency as observed for OTSUWW+NN and OTSUWW+U+NN and consequently missing edges were its main problem. With respect to false positive detections both TWANG-based methods provided the best results, with precision values of 1.0 due to a high quality seed detection that benefitted from the simultaneous multiview acquisition. This reflects an improvement of the precision obtained by the TWANG-based methods of 22.0% compared to OTSUWW+NN, which did not have an uncertainty-based object exclusion and thus an increased amount of false positive detections in background regions. The reason for the slightly higher number of redundant edges observed for the two TWANG-based methods is not yet fully clear. Most likely, the seed detection already provided a redundant seed to the segmentation method, which produced two nearby segments that were in turn counted as a redundant segment by the tracking evaluation. However, even for the

Table 5: Quantitative assessment of the tracking performance

Method	TP	FP	FN	Redundant	Missing	Merged	Rec.	Prec.	F-Score	TRA	Time (s)	KVoxel/s
OTSUWW+NN	905.33	194.10	110.67	5.51	256.88	94.33	0.89	0.82	0.86	0.81	51.36	1020.81
OTSUWW+U+NN	941.94	65.65	74.06	2.63	141.39	55.04	0.93	0.93	0.93	0.89	33.68	1556.68
TWANG+NN	889.73	1.73	126.27	13.75	191.16	3.29	0.88	1.00	0.93	0.86	14.27	3674.06
TWANG+U+NN	943.90	1.76	72.10	5.16	96.55	3.27	0.93	1.00	0.96	0.92	21.55	2432.89
OTSU+F+NN+U*	989.63	9.84	26.37	0.67	79.67	59.33	0.97	0.99	0.98	0.94	18.66	2809.72

Quantitative performance assessment of a nearest neighbor tracking algorithm (NN) applied on different segmentation results. Two algorithms without uncertainty-based improvements (OTSUWW, TWANG) were compared to enhanced pipelines that explicitly incorporated prior knowledge-based uncertainty treatment (OTSUWW+U, TWANG+U). Furthermore, an OTSU-based segmentation with additional seed points from LoGNSM+F+U (OTSU+F+NN+U) was used with an adapted tracking algorithm. Note that the segmentation produced by OTSU+F+NN+U was not usable for other purposes than tracking due to many merged regions (indicated by *). The validation measures correspond to true positives (TP), false positives (FP), false negatives (FN), redundant edges (Red.), missing edges (Miss.) and merged objects (Merg.). Furthermore, recall, precision, F-Score and the TRA measure were calculated as described in App. B and [32]. Processing times are average values for applying segmentation and tracking on a single image and were measured in seconds (lower values are better) and voxels per second (higher values are better).

worst algorithm in this category (TWANG+NN) redundant edges were only observed for 1.4% of the tracked objects and thus play a minor role compared to the other tracking errors. As shown in Fig. 9, the number of false negatives and merged objects directly correlated with the density of the objects, *i.e.*, the closer the objects are to each other, the more under-segmentation errors occurred. However, this was not the case for TWANG-based methods due to the explicit prior knowledge about the object size that is incorporated to the algorithm. The OTSU+F+NN+U method produced the best tracking result and was the second fastest method (TRA value of 0.94 and average processing time of 18.7s) closely followed by TWANG+U+NN (TRA value of 0.92 and average processing time of 21.6s). Compared to OTSUWW+NN, TWANG+U+NN and OTSU+F+NN+U provide superior quality in all categories and in particular an increase of the TRA measure by 13.6% and 16.0%, and a decrease of processing times by 58.0% and 63.7%, respectively. Thus, the two latter methods represent the best quality vs. speed trade-off and are suitable for large-scale analyses. Although OTSU+F+NN+U provided excellent results in this comparison, it should be noted that the extracted segmentation masks were largely merged and an object splitting approach as performed for OTSUWW would be required if object properties need to be known. An additional object splitting approach, however, would eradicate the performance benefit of the method and, *e.g.*, the TWANG+U+NN pipeline should be favored in this case.

To fall back on seed point information has no benefit for segmentation methods like TWANG, where the algorithmic design already only extracts a single segment per seed point and literally no merged objects exist. However, an interesting extension to consider in upcoming work might be a combination of the LoGNSM+F+U and the OTSU-based segmentation for seed detection and to feed these seeds to the TWANG algorithm, to reach both a further reduced amount of missed objects and a reduced amount of merged objects. Moreover, the temporal coherence was not yet considered in the investigated framework, *i.e.*, additionally allowing a nearest neighbor matching over multiple frames could potentially also help to reduce the number of missing and redundant detections.

3.4 Application to Light-Sheet Microscopy Images of Zebrafish Embryos

The presented framework was successfully used for the automated analysis of large-scale 3D+t microscopy images of developing zebrafish embryos [25, 46, 48, 49]. In particular, we used the LoGNSM+U method, *i.e.*, seed points were detected using a non-strict local maximum detection with a subsequent fusion of redundant detections and a false positive suppression based on the axial location of the seeds as well as their fluorescence intensity information (Sec. 2.2.1). These seeds were then provided to the TWANG algorithm as described in [48] and the segments of different views were combined using a segment-based fusion approach (App. C in Additional file 1 and [25]). Finally, a nearest-neighbor tracking was applied to the detected objects to obtain the movement trajectory-

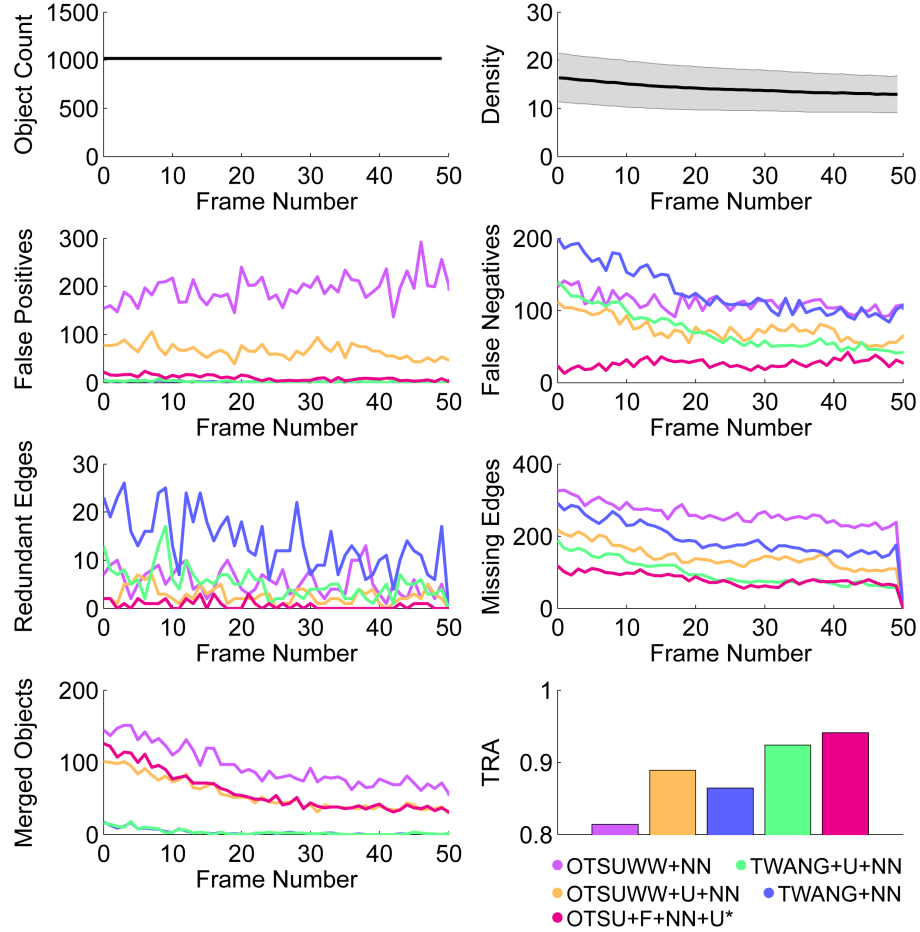


Figure 9: Quantitative performance assessment of a nearest neighbor tracking algorithm (NN) applied on different segmentation results obtained on the SBDE3 data set. Two algorithms without uncertainty-based improvements as described in [48] (OTSUWW, TWANG) were compared to enhanced pipelines that explicitly incorporate prior knowledge-based uncertainty treatment (OTSUWW+U, TWANG+U, OTSU+F+NN+U). (*) indicates that the respective algorithm did not produce a usable segmentation image, as the correction was solely performed at the tracking step.

ries (TWANG+U+NN). As there were about 8 million dynamically interacting objects in total for the investigated time period of about 3-10 hours post fertilization (hpf), manual labeling and validation were impossible. Qualitative results of the seed detection, the segmentation and the tracking stage are depicted in Fig. 10.

4 Conclusion

In this contribution, we presented a general concept for the mathematical formulation of prior knowledge and showed how image analysis pipelines can be equipped with the formalized prior knowledge to make more elaborate decisions. The framework includes the propagation of estimated result uncertainties, to be able to inform downstream pipeline operators about the validity of their input data and to potentially improve their results. Besides these general concepts, we demonstrated how an exemplary pipeline consisting of seed point detection, segmentation, multiview fusion and tracking could be systematically extended by the proposed uncertainty considerations, in order to filter, repair and fuse produced data. The performance of all proposed improvements was quantitatively assessed on a new and comprehensive validation benchmark inspired by light-sheet microscopy recordings of live specimen. The extensions proofed their superior performance compared to the plain pipelines and only had a low impact on the processing times due to the lightweight adaptations of the propagated feature matrices. Thus, the proposed framework represents a powerful approach to improve the quality and efficiency of image analysis pipelines. Several components of the presented framework were successfully used to analyze large-scale 3D+t light-sheet microscopy images of developing zebrafish embryos as described in Sec. 3.4 and [25, 44].

For simplicity, we used mostly focused on simple processing methods to illustrate the general concepts. However, extending more complex seed detection, segmentation or tracking algorithms with the presented concepts of filtering, splitting and fusion should work analogously if the uncertainty-based corrections are considered as a post-processing strategy of each processing operator. Especially the tracking step offers a lot of potential to be further improved. The uncertainty framework could be exploited to classify movement events, *e.g.*, to detect object divisions or to reconstruct missing objects using the temporal coherence of the objects. However, this was beyond the scope of this paper and will be addressed in upcoming work. In addition to the algorithmic improvements, we showed how the respective fuzzy sets can be parameterized based on available prior knowledge, such as feature histograms or knowledge about acquisition deficiencies. The respective shape of the fuzzy sets has to be determined based on the desired outcome. For instance, the false positive suppression at the seed point detection stage could be performed with a fixed threshold instead of an explicit usage of fuzzy sets for the intensity-based features. However, to model the increasing uncertainty (decreasing FSMD value) in regions farther away from the detection objective using a trapezoidal shape was more appropri-

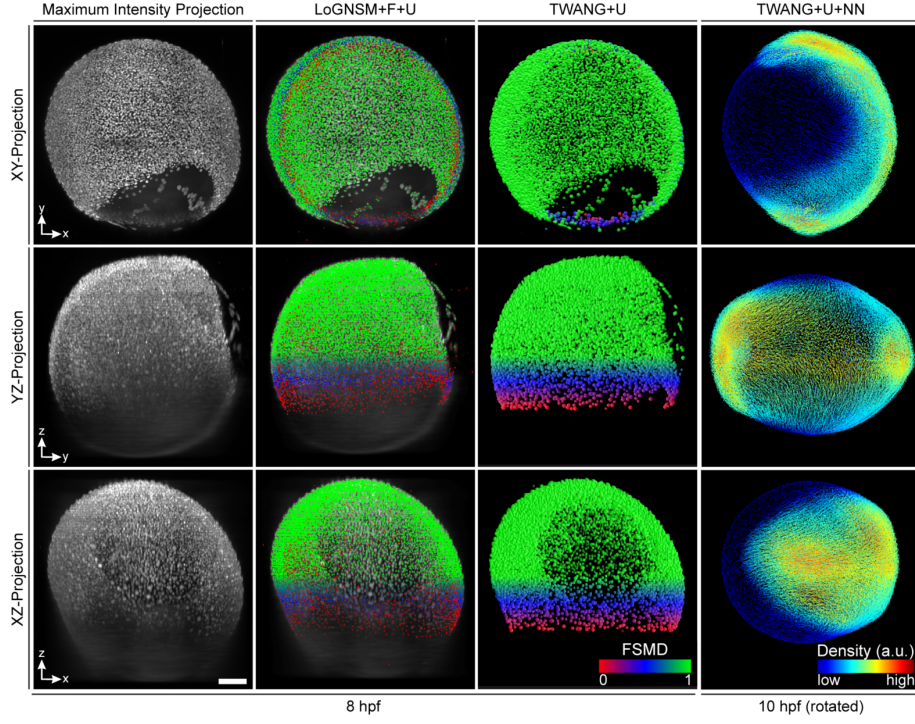


Figure 10: Application of the presented framework to 3D+t light-sheet microscopy images of zebrafish embryos. The panels show the maximum intensity projections on the XY, YZ and XZ planes. Seed points were extracted using the LoGNSM+F+U method (Sec. 2.2.1) and the improved seed points were used for segmentation using the TWANG+U method (Sec. 2.2.2). Extracted objects of two opposite views were combined using a segment-based fusion approach (App. C) and the fused object locations were then used for tracking (TWANG+U+NN, Sec. 2.2.3). The first three columns show results of a single view at approximately 8 hours post fertilization (hpf) with the color-code according to the FSMD values of the extracted objects (red over blue to green for low, medium and high FSMD values). The last column shows the results after multiview fusion and tracking at about 10 hpf using a cell density for coloring. In addition, the embryo was oriented along the coordinate axes, such that the anteroposterior axis was aligned with the Y axis (animal pole on the top) and the dorsoventral axis aligned with the X axis (dorsal to the right) [25]. Data were taken from our previously published work [25,46,48,49]. Scale bar, 100 μm .

ate. Further work has to be put on the automatic determination of the involved fuzzy sets, *e.g.*, using a semi-automatic approach for a manual classification of a representative subset of data.

Acknowledgements

The work was funded by the German Research Foundation DFG (JS, Grant No MI 1315/4, associated with SPP 1736 Algorithms for Big Data) and to the Helmholtz Association in the program BioInterfaces in Technology and Medicine (RM).

A Validation Benchmark

To provide a thorough validation of the entire pipeline comprised of seed point detection, segmentation, multiview fusion and tracking, we used our recently presented approach for generating comprehensive validation benchmarks [46] with an adapted object movement simulation. In brief, object locations, object movements and object interactions were simulated over multiple time points to obtain movement behaviors that resembled biological specimens. At each simulated position, a video snippet containing a simulated fluorescent cellular nucleus [51] was added to an artificial 3D image, to form the data basis for each time point. To simulate acquisition deficiencies, the simulated images were disrupted by additive Gaussian noise, Poisson shot noise, a point-spread-function simulation [38], light attenuation and multiview acquisition simulation as detailed in [46]. All acquisition deficiencies can be added to the simulated raw images using an XPIWIT pipeline [6] and the pipeline can be downloaded from <https://bitbucket.org/jstegmaier/embryomicsbenchmark/downloads/AcquisitionSimulationPipeline.zip>.

For the sake of simplicity, we simulated objects that were moving on a spherical surface. Instead of using object displacements of a real embryo as described in [46], the simulated objects only moved due to density changes and the resulting repulsive and adhesive forces acting between neighboring objects. Furthermore, the simulation was constrained to a spherical surface to prevent arbitrary movement in the simulation space. A schematic illustration of the simulated specimen and an overview of the involved simulation steps of the benchmark is shown in Fig. 11. The repulsive ($\Delta \mathbf{x}^{\text{rep}}$) and adhesive forces ($\Delta \mathbf{x}^{\text{adh}}$) as well as the parameterization was taken from [30] and the boundary constraint was defined as:

$$\Delta \mathbf{x}^{\text{bdr}}(\mathbf{x}, \mathbf{c}, r_i, r_o, a) = \begin{cases} \frac{\mathbf{x} - \mathbf{c}}{\|\mathbf{x} - \mathbf{c}\|} \cdot \left(1 - \frac{1}{e^{-a(\|\mathbf{x} - \mathbf{c}\| - r_i)}}\right), & \|\mathbf{x} - \mathbf{c}\| < r_i \\ -\frac{\mathbf{x} - \mathbf{c}}{\|\mathbf{x} - \mathbf{c}\|} \cdot \left(1 - \frac{1}{e^{a(\|\mathbf{x} - \mathbf{c}\| - r_o)}}\right), & \|\mathbf{x} - \mathbf{c}\| > r_o \\ \mathbf{0}, & \text{else.} \end{cases} \quad (3)$$

In Eq. (3), \mathbf{x} is the centroid of the considered object, \mathbf{c} is the center of the bounding volumes, r_i and r_o are the radii of the inner and the outer sphere, re-

spectively, and finally a controls the shape of the sigmoidal boundary potential function. $\Delta \mathbf{x}^{\text{bdr}}$ only contributed to the displacement of a simulated object if the object was already out of the boundary. This additional displacement component prevented objects from entering the inner bounding sphere and from escaping the outer bounding sphere of the simulated embryo (Fig. 11A). Analogous to the formulation used in [46], the total displacement vector of a single object at a given time point can be summarized to:

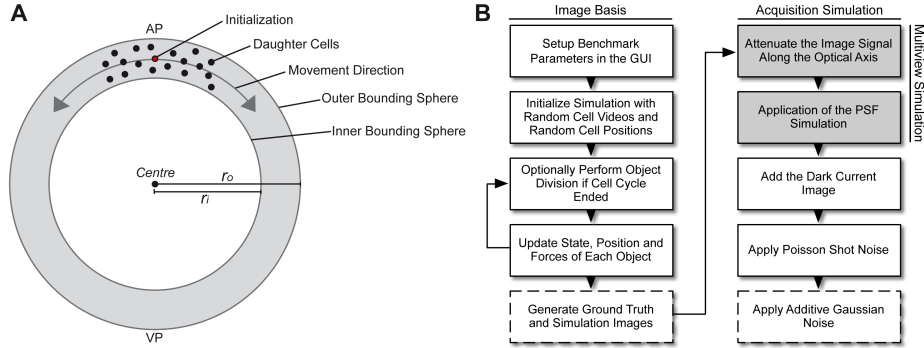


Figure 11: Pipeline schematic of the benchmark simulation. (A) Illustration of the embryo simulation. Starting with an initial object, multiple division cycles were simulated including object interaction, object divisions and morphological constraints. Due to inner and outer bounding spheres and the between-object interactions, cells migrated from the animal pole (AP) to the artificial vegetal pole (VP), which produced a behavior similar to the epiboly movement of zebrafish embryogenesis. (B) The performed steps for a realistic simulated 3D+t benchmark. The left column reflects the object simulation and returned raw images that contained dynamic objects and the associated ground truth data. The right column contains the acquisition simulation, which distorted the simulated images by an artificial signal attenuation, a point spread function simulation (PSF), a dark current image simulation, Poisson distributed photon shot noise and additive Gaussian noise. Steps shaded in gray could optionally consider image rotation, if a multiview experiment was simulated and the output operators are indicated by dashed edge lines (adapted from [46]).

$$\Delta \mathbf{x}_i^{\text{tot}} = w_{\text{bdr}} \cdot \Delta \mathbf{x}^{\text{bdr}}(\mathbf{x}_i) + \sum_{\substack{i \neq j \\ j \in \{1, \dots, N\}}} [w_{\text{rep}} \cdot \Delta \mathbf{x}^{\text{rep}}(\|\mathbf{x}_i - \mathbf{x}_j\|) + w_{\text{adh}} \cdot \Delta \mathbf{x}^{\text{adh}}(\|\mathbf{x}_i - \mathbf{x}_j\|)]. \quad (4)$$

The weights of the adhesive and repulsive displacement components were set to the default values mentioned in [30]: $w_{\text{adh}} = 0.52$, $w_{\text{rep}} = 1.0$. Moreover, the weight for the boundary constraint $w_{\text{bdr}} = 3.0$ was manually adjusted, such that

the interacting objects remained within the spherical boundaries. Note that these parameters were empirically determined and that the presented model does not necessarily represent an accurate physical simulation of the interacting objects. However, the determined parameters produced movement behaviors that were similar to the epiboly movements observed during early zebrafish development due to increased object densities and boundary constraints that caused a directed movement on a sphere surface. Exemplary benchmark images are shown in Fig. 12 for different time points and different additive Gaussian noise levels. Due to the availability of a complete ground truth, a detailed quantitative analysis of all involved pipeline steps could be performed with a single benchmark using the performance measures described in the following section. All validation datasets ($\approx 70\text{GB}$) used in this publication are available upon request.

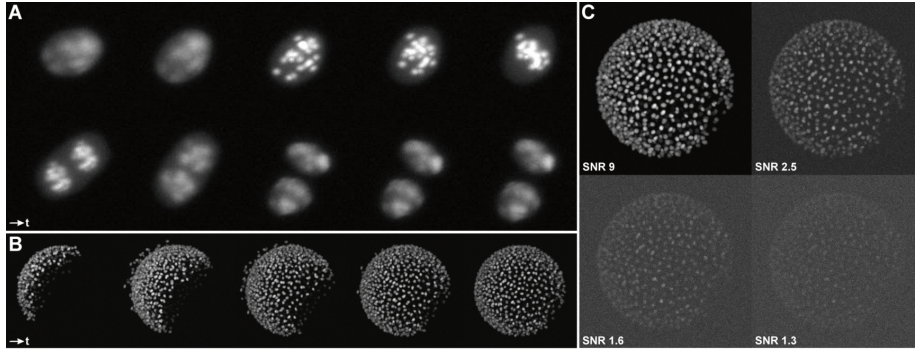


Figure 12: Maximum intensity projections of simulated benchmark images. (A) Maximum projections of an extracted division cycle of one simulated nucleus [51]. Time increases from left to right and top to bottom. Single objects were randomly initialized and simulated for a predefined experimental duration. This approach yielded a simulated embryo including object movement, object interaction and object divisions with available ground truth (B). Generated raw image sequences were manipulated to simulate various acquisition conditions, such as different levels of additive Gaussian noise (C). The ground truth enabled a quantitative analysis of the algorithmic performance on realistic image data.

B Performance Assessment

B.1 Seed Detection

The seed detection quality was evaluated using the benchmark datasets SBDE1 and SBDE2 (Tab. 6). The intersections of the detected seeds with the labeled ground truth image were calculated. True positives (TP) were counted as ground truth objects that contained at least one seed point. Seed points that were detected in background regions or redundant detections of ground truth objects

were considered as false positives (FP). Ground truth objects that did not contain a seed point were counted as false negatives (FN). Using TP, FP and FN, recall, precision and the F-Score (harmonic mean of precision and recall) were calculated. For all true positives, the average distance to the centroids of the respective ground truth objects was additionally calculated.

B.2 Segmentation

The segmentation quality was assessed using the benchmark datasets **SBDE1** and **SBDE2** (Tab. 6). As the provided ground truth of the benchmark contained the complete label images of each frame, a detailed quantitative assessment of the automatic segmentation quality could be performed. The set of segmentation validation measures proposed by Coelho *et al.* was used, namely the Rand index (RI), the Jaccard index (JI), the normalized sum of distances (NSD) and the Hausdorff metric (HM). A detailed description of the measures can be found in [18]. Topological errors produced by the automatic segmentation were separated into added, missing, split or merged objects. Besides the error counts, this topological information was used to define the number of false positives as the sum of split and added cells and analogously the false negatives as the sum of merged and missing cells. These values were then used to calculate recall, precision and F-Score.

B.3 Tracking

To assess the tracking quality, the **SBDE3** dataset was used (Tab. 6). The comparison of the investigated algorithms was performed using the **TRA** measure as described by Maška *et al.* [32]. This measure was calculated by considering the tracking result as an acyclic oriented graph and by comparing this graph to the respective ground truth graph. The inverted, weighted and normalized number of required changes to transform the automatically generated graph to the ground truth graph yielded the normalized **TRA** measure (higher values are better with 1 being ideal). As the centroids of all objects, the complete temporal association and the object ancestry was known for the simulated **SBDE3** dataset, the required ground truth graph could directly be generated using this data. To obtain a more detailed view on the errors made by the respective tracking algorithms, the number of false positive detections, false negative detections, incorrect edges, missing edges, redundant edges and merged objects were counted. Detailed descriptions of the validation measures are provided in [32].

B.4 Evaluation Platform

All measurements with respect to processing times were performed on a desktop PC equipped with an Intel Core i7-2600 CPU @ 3.4GHz and 32GB of memory installed using the Windows 7 (x64) operating system.

C Multiview Fusion

A frequently used approach to partly overcome quality deficiencies in 3D microscopy that are caused by attenuation and scattering of light along the axial direction is the acquisition of multiple views from different perspectives [17, 26]. Such complementary image stacks of a specimen can be obtained either by using multiple oppositely arranged detection paths [17, 53] or by a rotation of the probe using a single camera [25]. Both strategies for the acquisition of multiview images require a subsequent fusion step of the information present in the individual views into a single consistent representation. The benchmark dataset SBDE3 used for the tracking validation contains such a simulated multiview acquisition, *i.e.*, at each time point, two opposite images were simulated. To fuse to complementary views, we use a segmentation-based fusion approach that combines the results of separately applied segmentation on the different views to a single segmentation image. As the image transformation of 180° was already known from the benchmark generation step, the registration process was skipped. Corresponding segments were identified using a histogram-based approach, *i.e.*, a 2D label histogram was filled by iterating over all voxels of both 3D segmentation images and by successively increasing the 2D histogram bins indicated by the respective image label pairs. Segments that were only present in one or the other image could be found by searching for empty columns and rows of the histogram, respectively, and by copying these segments without any further consideration into the new result image. In the next step, the assignments of the remaining segments in both images were identified by searching for the respective label pair with the maximum overlap. In the case of segmentation fusion, it was not desired to perform a weighted average of the segments but rather to use the better segment for the new image. This was accomplished by simply selecting the segment with a higher FSMD value to the desired class of objects based on the FSMD values of the segments and the seed points that contained the information about the validity of the segment and the axial localization of the extracted objects, respectively. The final segmentation image was then used for tracking.

D Availability of Data and Material

All validation data sets used in this publication (≈ 70 GB) are made available upon request. MATLAB scripts for generating benchmark data sets of simulated embryos as well as the XPIWIT pipeline that was used for acquisition simulation can be found at <https://bitbucket.org/jstegmaier/embryomicsbenchmark/>. Furthermore, source code as well as binary packages of XPIWIT are available from <https://bitbucket.org/jstegmaier/xpiwit/>.

Table 6: The benchmark datasets used for the validation experiments. Columns list the dataset identifier (Name), the number of images (Im.), the number of noise levels (N.), the number of objects (Objects), the image resolution, the noise parameter ranges for σ_{agn} and the dataset description.

Name	Im.	N.	Objects	Resolution	σ_{agn}	Description
SBDE1	10	1	300–1000	$640 \times 640 \times 128$	0.0007	Five pairs of subsequent frames with different numbers of objects. Used for seed detection and segmentation validation.
SBDE2	20	20	1000	$640 \times 640 \times 128$	[0.0005, 0.01]	Single frame of SBDE1 with 20 different noise levels. Used for seed detection and segmentation validation.
SBDE3	100	1	1000	$640 \times 640 \times 128$	0.0007	Sequential time points containing 1000 objects each. Images are rotated by 180° at every frame to simulate a simultaneous acquisition multiview experiment SiMV (2×50 images). Used for the validation of the tracking.

Table 7: Abbreviations, parameterizations and descriptions of the investigated seed detection algorithms.

Method	Parameters	Description
LoGSM	$\sigma_{\min} = 4, \sigma_{\max} = 8, \sigma_{\text{step}} = 1, t_{\text{wmi}} = 0.0025$	Seed detection in the LoG scale space maximum projection with a manually adjusted window mean intensity threshold (t_{wmi}) and a strict maximum detection.
LoGNSM	$\sigma_{\min} = 4, \sigma_{\max} = 8, \sigma_{\text{step}} = 1, t_{\text{wmi}} = 0.0025$	Seed detection in the LoG scale space maximum projection with a manually adjusted window mean intensity threshold (t_{wmi}) and a non-strict maximum detection.
LoGNSM+F	$\sigma_{\min} = 4, \sigma_{\max} = 8, \sigma_{\text{step}} = 1, t_{\text{wmi}} = 0.0025, t_{\text{dbc}} = 5$	Same detection as LoGNSM but with additional fusion (F) of redundant detections using a hierarchical clustering approach with a distance-based cutoff value (t_{dbc}).
LoGNSM+F+U	$\sigma_{\min} = 4, \sigma_{\max} = 8, \sigma_{\text{step}} = 1, t_{\text{dbc}} = 5, \boldsymbol{\theta}_{\text{wmi}} = (0.0025, 0.0025, \infty, \infty), \boldsymbol{\theta}_{\text{smi}} = (0.0007, 0.0007, \infty, \infty), \boldsymbol{\theta}_{\text{zpos}} = (50, 250, \infty, \infty), \alpha_{11} = 0.0001, \beta_{11} = \alpha_{11}$	Same detection as LoGNSM but with uncertainty-based (U) threshold and the fusion of LoGNSM+F. The forward threshold α_{11} is set slightly above zero, such that obvious false positives are rejected. As no further processing was needed β_{11} was set to α_{11} .

Table 8: Abbreviations, parameters and descriptions of the segmentation algorithms.

Method	Parameters	Description
OTSU	$\sigma_{\text{smooth}} = 1$	Adaptive thresholding using Otsu's method applied on a regularized raw input image [35].
OTSUWW	$\sigma_{\text{smooth}} = 1$	Same segmentation as OTSU. Merged regions in the entire image were split based on cleaned seeds of the LoGSM method, using a seeded watershed approach.
OTSUWW+U	$\sigma_{\text{smooth}} = 1$, $\alpha_{21} = 0.1$, $\beta_{21} = 0.1$	Same segmentation as OTSU. Small noise objects with FSMD values below $\alpha_{21} = 0.1$ were rejected from further processing. Merged regions with FSMD below $\beta_{21} = 0.1$ that were larger than expected were split locally and in parallel using the cleaned and fused seeds of the LoGNSM+F+U method and a seeded watershed approach in small cropped regions.
TWANG	$\sigma_{\text{grad}} = 3$, $\sigma_{\text{kernel}} = 3$, $\omega_{\text{kpm}} = 1.41$	TWANG segmentation as described in [48] using the cleaned seeds provided by the LoGSM method. The manually optimized parameter $\omega_{\text{kpm}} = 1.41$ yielded better results than the default value of 1.0 for this dataset.
TWANG+U	$\sigma_{\text{grad}} = 3$, $\sigma_{\text{kernel}} = 3$, $\omega_{\text{kpm}} = 1.41$, $\alpha_{21} = 0$, $\beta_{21} = 0$	TWANG segmentation as described in [48] using the cleaned and fused seeds provided by the LoGNSM+F+U method. $\alpha_{21} = 0$, $\beta_{21} = 0$ are adjusted such that all segments are propagated unchanged.

References

- [1] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering*, 57(4):841–852, 2010.
- [2] F. Amat, W. Lemon, D. P. Mossing, K. McDole, Y. Wan, K. Branson, E. W. Myers, and P. J. Keller. Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. *Nature Methods*, 11(9):951–958, 2014.
- [3] R. Anchini, C. Liguori, and A. Paolillo. Evaluation of the uncertainty of edge-detector algorithms. *IEEE Transactions on Instrumentation and Measurement*, 56(3):681–688, 2007.

- [4] D. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.
- [5] Z. Bao, J. Murray, T. Boyle, S. Ooi, M. Sandel, and R. Waterston. Automated cell lineage tracing in *caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America*, 103(8):2707–2712, 2006.
- [6] A. Bartschat, E. Hübner, M. Reischl, R. Mikut, and J. Stegmaier. XPI-WIT - an XML pipeline wrapper for the insight toolkit. *Bioinformatics*, 32(2):315–317, 2016.
- [7] R. Beare and G. Lehmann. The watershed transform in ITK-discussion and new developments. *The Insight Journal*, pages 1–24, 2006.
- [8] B. Bede. *Mathematics of Fuzzy Sets and Fuzzy Logic*. Springer, Berlin Heidelberg, 2013.
- [9] G. Betta, D. Capriglione, M. Corvino, C. Liguori, and A. Paolillo. Estimation of influence quantities in face recognition. In *IEEE International Instrumentation and Measurement Technology Conference*, pages 963–968. IEEE, 2012.
- [10] G. Betta, D. Capriglione, C. Liguori, and A. Paolillo. Uncertainty evaluation in face recognition algorithms. In *IEEE International Instrumentation and Measurement Technology Conference*, pages 1–6, may 2011.
- [11] V. Boskovitz and H. Guterman. An adaptive neuro-fuzzy system for automatic image segmentation and edge detection. *IEEE Transactions on Fuzzy Systems*, 10(2):247–262, 2002.
- [12] B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh. *Information, Uncertainty and Fusion*. Kluwer Academic, Boston, 2000.
- [13] P. Bourguine, R. Čunderlík, O. Drblíková-Stašová, K. Mikula, M. Remešíková, N. Peyriéras, B. Rizzi, and A. Sarti. 4D embryogenesis image analysis using pde methods of image processing. *Kybernetika*, 46(2):226–259, 2010.
- [14] B. Callejas Bedregal, G. P. Dimuro, and A. Rocha Costa. Interval fuzzy rule-based hand gesture recognition. In *12th GAMM-IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics*, pages 12–12. IEEE, 2006.
- [15] J. Chen and L. Zhao. A new method for uncertainty evaluation of centroid detection. In *2nd International Congress on Image and Signal Processing, 2009. CISP’09*, pages 1–4. IEEE, 2009.
- [16] J. Chen and L. Zhao. A new method for uncertainty evaluation of corner detection. In *International Conference on Artificial Intelligence and Computational Intelligence*, volume 1, pages 458–462. IEEE, 2010.

- [17] R. K. Chhetri, F. Amat, Y. Wan, B. Höckendorf, W. C. Lemon, and P. J. Keller. Whole-animal functional and developmental imaging with isotropic spatial resolution. *Nature Methods*, 12(12):1171–1178, 2015.
- [18] L. Coelho, A. Shariff, and R. Murphy. Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms. In *Proc., IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 518–521, 2009.
- [19] F. de Chaumont, S. Dallongeville, N. Chenouard, N. Hervé, S. Pop, T. Provoost, V. Meas-Yedid, P. Pankajakshan, T. Lecomte, Y. Le Montagner, T. Lagache, A. Dufour, and J. C. Olivo-Marin. Icy: An open bioimage informatics platform for extended reproducible research. *Nature Methods*, 9(7):690–696, 2012.
- [20] E. Faure, T. Savy, B. Rizzi, C. Melani, O. Stašová, D. Fabrèges, R. Špir, M. Hammons, R. Čúnderlík, G. Recher, et al. A workflow to process 3D+time microscopy images of developing organisms and reconstruct their cell lineage. *Nature Communications*, 7(8674):1–10, 2016.
- [21] R. Fernandez, P. Das, V. Mirabet, E. Moscardi, J. Traas, J.-L. Verdeil, G. Malandain, and C. Godin. Imaging plant growth in 4D: Robust tissue reconstruction and lineaging at cell resolution. *Nature Methods*, 7(7):547–553, 2010.
- [22] A. Khan, R. Mikut, B. Schweitzer, C. Weiss, and M. Reischl. Automatic tuning of image segmentation parameters by means of fuzzy feature evaluation. *Advances in Intelligent Systems and Computing*, 190(5):459–467, 2013.
- [23] A. Khan, M. Reischl, B. Schweitzer, C. Weiss, and R. Mikut. Feedback-driven design of normalization techniques for biological images using fuzzy formulation of a priori knowledge. *Studies in Computational Intelligence*, 445:167–178, 2013.
- [24] Z. Khan, Y.-C. Wang, E. F. Wieschaus, and M. Kaschube. Quantitative 4D analyses of epithelial folding during drosophila gastrulation. *Development*, 141(14):2895–2900, 2014.
- [25] A. Kobitski, J. C. Otte, M. Takamiya, B. Schäfer, J. Mertes, J. Stegmaier, S. Rastegar, F. Rindone, V. Hartmann, R. Stotzka, A. García, J. v. Wezel, R. Mikut, U. Strähle, and G. U. Nienhaus. An ensemble-averaged, cell density-based digital model of zebrafish embryo development derived from light-sheet microscopy data with single-cell resolution. *Scientific Reports*, 5(8601):1–10, 2015.
- [26] U. Krzic, S. Gunther, T. Saunders, S. Streichan, and L. Hufnagel. Multiview light-sheet microscope for rapid *in toto* imaging. *Nature Methods*, 9(7):730–733, 2012.

- [27] J. Kybic. Bootstrap resampling for image registration uncertainty estimation without ground truth. *IEEE Transactions on Image Processing*, 19(1):64–73, 2010.
- [28] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings.*, volume 1, pages 316–323. IEEE, 2000.
- [29] X. Lou, F. Kaster, M. Lindner, B. Kausler, U. Kothe, B. Hockendorf, J. Wittbrodt, H. Janicke, and F. Hamprecht. Deltr: Digital embryo lineage tree reconstructor. In *Proc., IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1557–1560. IEEE, 2011.
- [30] P. Macklin, M. E. Edgerton, A. M. Thompson, and V. Cristini. Patient-calibrated agent-based modelling of ductal carcinoma in situ (dcis): From microscopic measurements to macroscopic predictions of clinical progression. *Journal of Theoretical Biology*, 301:122–140, 2012.
- [31] P. Maji and S. K. Pal. *Rough-Fuzzy Pattern Recognition: Applications in Bioinformatics and Medical Imaging*, volume 3. John Wiley & Sons, Hoboken, New Jersey, 2011.
- [32] M. Maška, V. Ulman, D. Svoboda, P. Matula, P. Matula, C. Ederra, A. Urbiola, T. España, S. Venkatesan, D. M. Balak, et al. A benchmark for comparison of cell tracking algorithms. *Bioinformatics*, 30(11):1609–1617, 2014.
- [33] A. Mencattini, G. Rabottino, S. Salicone, and M. Salmeri. Uncertainty modeling and propagation through RFVs for the assessment of CADx systems in digital mammography. *IEEE Transactions on Instrumentation and Measurement*, 59(1):27–38, 2010.
- [34] D. B. Murphy and M. W. Davidson. *Fundamentals of Light Microscopy and Electronic Imaging*. Wiley-Blackwell, Hoboken, New Jersey, 2012.
- [35] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, 1979.
- [36] S. K. Pal. Fuzzy image processing and recognition: Uncertainty handling and applications. *International Journal of Image and Graphics*, 1(02):169–195, 2001.
- [37] H. Peng, A. Bria, Z. Zhou, G. Iannello, and F. Long. Extensible visualization and analysis for multidimensional images using Vaa3D. *Nature Protocols*, 9(1):193–208, 2014.
- [38] S. Preibisch, F. Amat, E. Stamatakis, M. Sarov, R. H. Singer, E. Myers, and P. Tomancak. Efficient bayesian-based multiview deconvolution. *Nature Methods*, 11(6):645–648, 2014.

- [39] M. Radojevic, I. Smal, and E. Meijering. Automated neuron morphology reconstruction using fuzzy-logic detection and Bayesian tracing algorithms. In *Proc., IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 885–888, 2015.
- [40] M. Santo, C. Liguori, A. Paolillo, and A. Pietrosanto. Standard uncertainty evaluation in image-based measurements. *Measurement*, 36(3):347–358, 2004.
- [41] M. Schiegg, P. Hanslovsky, B. X. Kausler, L. Hufnagel, and F. A. Hamprecht. Conservation tracking. In *IEEE International Conference on Computer Vision*, pages 2928–2935. IEEE, 2013.
- [42] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682, 2012.
- [43] J. Stegmaier. *New Methods to Improve Large-Scale Microscopy Image Analysis with Prior Knowledge and Uncertainty (in preparation)*. PhD thesis, Karlsruhe Institute of Technology, 2016.
- [44] J. Stegmaier, R. Alshut, M. Reischl, and R. Mikut. Information fusion of image analysis, video object tracking, and data mining of biological images using the open source MATLAB toolbox Gait-CAD. *Biomedizinische Technik (Biomedical Engineering)*, 57 (S1):458–461, 2012.
- [45] J. Stegmaier, F. Amat, W. B. Lemon, K. McDole, Y. Wan, G. Teodoro, R. Mikut, and P. J. Keller. Real-time three-dimensional cell segmentation in large-scale microscopy data of developing embryos. *Developmental Cell*, 36(2):225–240, 2016.
- [46] J. Stegmaier, J. Arz, B. Schott, J. C. Otte, A. Kobitski, G. U. Nienhaus, U. Strähle, P. Sanders, and R. Mikut. Generating semi-synthetic validation benchmarks for embryomics. In *Proc., IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2016.
- [47] J. Stegmaier, A. Khan, M. Reischl, and R. Mikut. Challenges of uncertainty propagation in image analysis. In *Proc., 22. Workshop Computational Intelligence, Dortmund*, pages 55–69, 2012.
- [48] J. Stegmaier, J. C. Otte, A. Kobitski, A. Bartschat, A. Garcia, G. U. Nienhaus, U. Strähle, and R. Mikut. Fast segmentation of stained nuclei in terabyte-scale, time resolved 3D microscopy image stacks. *PLoS ONE*, 9(2):e90036, 2014.
- [49] J. Stegmaier, B. Schott, E. Hübner, M. Traub, M. Shahid, M. Takamiya, A. Kobitski, V. Hartmann, R. Stotzka, J. van Wezel, A. Streit, G. U.

- Nienhaus, U. Strähle, M. Reischl, and R. Mikut. Automation strategies for large-scale 3D image analysis. *at - Automatisierungstechnik*, 64(7):555–566, 2016.
- [50] C. Straehle, U. Koethe, G. Knott, K. Briggman, W. Denk, and F. A. Hamprecht. Seeded watershed cut uncertainty estimators for guided interactive segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 765–772. IEEE, 2012.
 - [51] D. Svoboda and V. Ulman. Generation of synthetic image datasets for time-lapse fluorescence microscopy. In *Image Analysis and Recognition*, pages 473–482. Springer, 2012.
 - [52] H. R. Tizhoosh. Image thresholding using type II fuzzy sets. *Pattern Recognition*, 38(12):2363–2372, 2005.
 - [53] R. Tomer, K. Khairy, F. Amat, and P. Keller. Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy. *Nature Methods*, pages 755–763, 2012.
 - [54] N. Vu and B. Manjunath. Shape prior segmentation of multiple objects with graph cuts. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.
 - [55] J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.
 - [56] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.
 - [57] X. Zhou, A. Gupta, and D. Comaniciu. An information fusion framework for robust shape tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):115–129, 2005.